AD-A239 011

(2)

UNCLASSIFIED

TECHNICAL NOTES ON

RISK AND UNCERTAINTY IN EARLY COMPARABILITY ANALYSIS

JANUARY 1990

TRADOC ANALYSIS COMMAND - FORT BENJAMIN HARRISON

UNCLASSIFIED

91 7 26 034

91-06240

UNCLASSIFIED

RISK AND UNCERTAINTY IN EARLY COMPARABILITY ANALYSIS

by

Gerald A. Klopp, Ph.D.

January 1990

Director
US Army TRADOC Analysis Command
ATTN: ATRC-B
Fort Benjamin Harrison, Indiana 46216-5000

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188 Exp Date Jun 30 1986

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) TRAC-FBHN	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION US Army TRADOC Analysis Command 6b. OFFICE SYMBOL (if applicable) ATRC-B	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) ATTN: ATRC-B (Bldg 401B) Fort Benjamin Harrison, IN 46216-5000	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION HQ, TRADOC, Combat Service Support Directorate 8b. OFFICE SYMBOL (if applicable) ATRC-SP	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Fort Monroe, VA 23651-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)

Risk and Uncertainty In Early Comparability Analysis

12. PERSONAL AUTHOR(S)

Gerald A. Klopp

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 1/90 TO 1/90	90/1/31	31

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	MANPRINT, Early Comparability Analysis (ECA), Risk, Uncertainty, HARDMAN

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In the early phases of the materiel acquisition process, Subject Matter Experts (SMEs) may be the only source for data on a conceptual weapon system. In these early phases, the Early Comparability Analysis (ECA) Methodology provides a tool for identifying the Manpower, Personnel, and Training (MPT) resource intensive tasks (e.g., the "high drivers") on currently fielded systems that the conceptual system will replace. Various documents related to Manpower and Personnel Integration (MANPRINT) and the Concept Formulation Process (CFP) discuss risk and uncertainty. By expanding these concepts and using some theoretical considerations, the ECA method can be extended to provide a probabilistic estimate of the uncertainty of the SMEs opinion and the risk of the proposed system. The process which determines the risk and uncertainty probabilities can then be worked "backward" to determine the high drivers as well as the source for risk and uncertainty. Hypothetical data from a very small sample of SMEs is used to illustrate the risk/uncertainty process. The method incorporates a consideration for "importance" values or weights in the determination of risk/uncertainty values probabilities.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL GERALD A. KLOPP, Director	22b. TELEPHONE (Include Area Code) (317) 543-6899 (AV 699) 22c. OFFICE SYMBOL ATRC-B

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

SPECIAL TECHNICAL NOTICE

This paper is an unoffical document intended for wide distribution to obtain comments. The views, opinions, and/or findings contained herein are those of the author and should not be construed as the official position of the TRADOC Analysis Command, US Army Training and Doctrine Command, or an official Department of the Army position, policy, or decision unless so designated by other official documentation.

Comments should be sent to:

Director, US Army TRADOC Analysis Command - Ft. Harrison
ATTN: ATRC-B (Bldg 401B)
Ft. Benjamin Harrison, IN 46216-5000

NOTICES

DESTRUCTION NOTICE

When this report is no longer needed, Department of the Army organizations will destroy it in accordance with procedures given in AR 380-5. Navy and Air Force elements will destroy it in accordance with applicable directives. Department of Defense contractors will destroy the report according to their requirements of Section 14 of the Industrial Security Manual for Safeguarding Classified Information. All others will return the report to Headquarters, United States Army Training and Doctrine Command.

DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

DISCLAIMER

The finding(s) of this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents. The words 'he', 'his', 'man' and 'men', when used in this publication, represent both the masculine and feminine genders unless otherwise specifically stated.

SECURITY CHECKLIST

1. TITLE OF STUDY: Risk and Uncertainty in Early Comparability Analysis.

2. This report is unclassified.

3. Limitations on dissemination have not been imposed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF ACRONYMS

AR	Army Regulation
ARI	Army Research Institute
CFP	Concept Formulation Process
DR	Decay Rate
ECA	Early Comparability Analysis
FR	Frequency Rate
MANPRINT	Manpower and Personnel Integration
MOI	Memorandum of Instruction
MOS	Military Occupational Specialties
MPT	Manpower, Personnel, and Training
MRA	MANPRINT Risk Assessment
PP	Percent Performing Task
SME	Subject Matter Expert
SMEs	Subject Matter Experts
SMMP	System MANPRINT Management Plan
SSC-NCR	Soldier Support Center-National Capital Region
TLD	Task Learning Difficulty
TPD	Task Performance Difficulty
TT	Time to Train
USAPIC	US Army Personnel Integration Command
WT_i	Weight

ABSTRACT

In the early phases of the materiel acquisition process, Subject Matter Experts (SMEs) may be the only source for data on a conceptual weapon system. In these early phases, the Early Comparability Analysis (ECA) Methodology provides a tool for identifying the Manpower, Personnel, and Training (MPT) resource intensive tasks (e.g., the 'high drivers') on currently fielded systems that the conceptual system will replace. Various documents related to Manpower and Personnel Integration (MANPRINT) and the Concept Formulation Process (CFP) discuss risk and uncertainty. By expanding these concepts and using some theoretical considerations, the ECA method can be extended to provide a probabilistic estimate of the uncertainty of the SME's opinion and the risk of the proposed system. The process which determines the risk and uncertainty probabilities can then be worked 'backward' to determine the high drivers as well as the source for risk and uncertainty. Hypothetical data from a very small sample of SMEs is used to illustrate the risk/uncertainty process. The method incorporates a consideration for 'importance' values or weights in the determination of risk/uncertainty values probabilities.

RISK AND UNCERTAINTY IN EARLY COMPARABILITY ANALYSIS

1. BACKGROUND:

According to AR 602-2, Manpower and Personnel Integration (MANPRINT) in the Materiel Acquisition Process, the purpose of the MANPRINT philosophy is to provide a systematic approach to gathering information to answer the question: 'can this soldier, with this training, perform these tasks to these standards under these conditions?' In the early phases of the materiel acquisition process, there may be a lack of engineering or other 'hard' data on the proposed system, so a 'comparability analysis' may be required. Data are usually obtained from Subject Matter Experts (SMEs) as well as other sources (such as sample data collection on the predecessor system which is to be replaced). Typically, these front-end analyses which are performed during the pre-concept phases of the system design will focus on the predecessor system and on lessons learned.

The Early Comparability Analysis (ECA) methodology, according to the ECA Handbook (Reference b(2)), is a 'lessons learned' approach to identify the Manpower, Personnel, and Training (MPT) resource intensive tasks ('high drivers') on currently fielded materiel systems that must be resolved in new or product improved systems. Thus, a major focus of the ECA methodology is on the 'task.' This handbook illustrates a procedure for identifying high drivers by using SMEs and other data and some easy arithmetic calculations.

The draft MANPRINT Analysis Methodology (Reference b(3)) shows that ECA is a part of the pre-concept, pre-milestone zero, phase of MANPRINT activities and is one of several processes which 'feed' the development of the System MANPRINT Management Plan (SMMP). As issues arise, the SMMP documents them. As the issues are addressed, the SMMP is updated, thus resulting in a 'living document' which changes as the state of knowledge changes for the proposed new system.

To get an indication of uncertainty, we look at the Concept Formulation Process (CFP) Memorandum of Instruction (MOI) (See Reference b(5)). The purpose of an uncertainty analysis is 'to put a range of probabilities around the estimate at a specific level of confidence; i.e., to bound the estimate.' Thus, uncertainty involves the estimation of a value and the likely or probable range of values that can occur. When we define a critical value for some system performance measurement, any value which exceeds that critical value is an indication of an unacceptable condition. Thus, if there is a range of likely or probable values which exceed the critical value, the range of all such values is an indication of the risk probability.

The MANPRINT Risk Assessment (MRA) Guide (Reference b(1)), contains a procedural 'tool to evaluate MANPRINT risk associated with the development of an emerging materiel system.' The user of this guide can obtain an indication of the 'risk' of a conceptual system by answering a series of Yes-No questions. As in the MRA, the risk methodology in this report is applicable to systems in the early concept phase. Both methods can be used to identify the need for further analyses or voids in understanding. However, the methodology in this report will also provide a probability estimate of the degree of risk associated with the system, a specific task, and the components of each task.

We will look at uncertainty and risk from an ECA standpoint. From the previous observations, we see that the discussion will focus on tasks and a set of criteria for evaluating task performance. Risk and uncertainty will be evaluated with respect to estimated values of task performance which exceed some critical level of overall acceptable performance. An example will be given using judgements from five SMEs, but the methodology is appropriate for much larger samples of SMEs. Indeed, statistical results are much more reliable when larger samples (in the range of 20 to 30 SME) are used.

Although the procedures discussed herein use the initial SME results, there are techniques (such as the Delphi Method) for allowing the 'experts' to use additional information to 'refine' their opinion (and hence, estimates). The object of the risk-uncertainty analysis should be to use all of the information available and not to exclude someone's opinion simply because it is not the same as other estimates.

2. INTRODUCTION.

We will use the data in Chapter 4 of (draft) TRADOC Pam 602-1 (Early Comparability Analysis Procedural Guide) to illustrate several features of an ECA which can provide additional insight to decision makers. Although the ECA manual strongly suggests a minimum sample of ten SME, the procedures are illustrated below and in the ECA manual using only five SMEs. First, let us review the ECA methodology, the data, and how it was collected because these are important considerations in the validity of the insights which we will subsequently address.

The ECA methodology depends heavily on soliciting information from Subject Matter Experts (SMEs). Most often, the SMEs will be soldiers in the Military Occupational Specialties (MOS) that either operate or maintain the predecessor system. In the very early phases of the system acquisition process, ECA may be applied to the conceptual system. However, SME opinion may be the only way for collecting data for an ECA on a conceptual system and is a very appropriate source if the SMEs are truly 'experts.' Step 3 of the ECA methodology calls for compiling a list of tasks. Step 4 calls for collecting data from the SMEs using a well-defined structured approach whereby the SMEs rate each task on each of six criteria.

For an ECA for which there is a predecessor system, a myriad of data in addition to SME opinion should also be available. The six criteria are rated on a variable scale, with possible values between one and four. A score of 'one' usually is a 'desirable' rating (e.g., low percentage of MOS performing the task, task not difficult, task is seldom performed, low task proficiency decay, or small number of hours to train) and a score of 'four' is an 'undesirable' rating (e.g., high percentage performing the task, task is difficult, task frequently performed, high proficiency decay, or large number of hours to train). These 'undesirable' characteristics may result in a task that is referred to as a 'high driver' of personnel and cost. The six task rating criteria are listed in Table 1, and the sample data from five SMEs for the six criteria are listed in Table 2.

Table 1. ECA Task Rating Criteria

Criteria Abbreviation	Criteria Description
PP	Percent Performing Task
TLD	Task Learning Difficulty
TPD	Task Performance Difficulty
FR	Frequency Rate
DR	Decay Rate
TT	Time to Train

Table 2. SME Evaluations of One Task

SME	PP	TLD	TPD	FR	DR	TT
1	4	1	1	3	1	2
2	3	1	1	3	1	3
3	3	2	2	4	2	2
4	3	2	1	4	3	1
5	2	1	2	3	3	3
Average	3.0	1.4	1.4	3.4	2.0	2.2

The ECA methodology results in the calculation of a 'Task Score' by computing the product of the averages of the SME scores. Using the data above, the Task Score is:

$$3.0 \times 1.4 \times 1.4 \times 3.4 \times 2.0 \times 2.2 = 87.96.$$

Using the ECA methodology, this Task Score is compared to an 'established' cutoff score of 216. Since the Task Score is less than 216, the task is not considered a 'High Driver.' The procedure summarized above would be completed for each task to identify the 'High Drivers' of the system being evaluated.

3. UNCERTAINTY AND RISK.

Whereas the ECA method provides useful information on currently fielded systems, as we will see below, with modifications, it can provide useful information on conceptual systems as well. The modifications discussed below will enhance the basic methodology for evaluating either conceptual or currently fielded systems. First, note in Table 2, the SMEs have some disagreement in their assessments of each of the six criteria. The ECA methodology seeks to form a concensus by using the average value for each of the six criteria. However, the variation of the SME evaluations is an indication of their uncertainty. We could also argue that high variation in SME opinion may be an indication of the likelihood that problems will be encountered in the concept formulation phase of the system acquisition process. Certainly, variation in SME opinion translates into uncertainty.

To illustrate the need for further ECA methodology enhancements, suppose that we could get another 'set' of five SME responses. On the basis of the first sample of five, we should conclude that the second sample would probably also exhibit some uncertainty. Of course, to gain 'perfect' information on the uncertainty of the SMEs (e.g., to determine their variation in task criteria scores), we would have to gather every SME's opinion. However, if we are willing to consider the first sample as 'representative' of any and all samples of SME opinion, we can use probability and statistics to make inferences about the other possible samples of SME opinion.

If we carry the considerations on the SME sample further, with some assumptions about the relationship between the six criteria, we can use the SME uncertainty to provide us information on the 'risk' of the task as well. First, let us assume that the SMEs evaluated each of the six criteria separately (that is, for example, criterion one was not influenced by nor did it influence the rating given to criterion two). We refer to this property as independence. This assumption was implicitly built into the ECA methodology when the Task Score was computed using the product of the average criteria values. Although a desirable property, as discussed in the Post Script (Section 9), independence is not necessary for the ECA methodology.

We can consider the SME criteria scores to be taken from a distribution of scores. As we noted earlier, we cannot practically collect all SME scores to gain 'perfect' information, so we will use this sample of SMEs to give us an idea of the distribution characteristics. That is, we can use the sample characteristics to estimate how the entire SME population would have scored each independent criterion. First, we will construct a distribution of possible criterion scores using the SME data to estimate key parameters which describe the distributions. Next, we will construct a distribution of Task Scores by randomly drawing values from the sample criterion distributions. The procedure for constructing the Task Score distribution will be illustrated later. We will now discuss the characteristics of the criterion distributions.

There are several theoretically sound distributions to use. The 'easiest' distribution would be a normal distribution whose average and variance (or standard deviation) is estimated by the sample data average and variance (or standard deviation). However, we have to place some constraint on the distribution of possible values. In a normal distribution, it is possible (but not very probable) to have a score several standard deviations below the average (or mean). This could equate to a meaningless score of '0' or a negative criterion value. Since the sample of SMEs could not provide any score other than a value between 1 and 4, we must 'truncate' the theoretical normal distribution so that only values between 1 and 4 are possible. Of course, we could open up the range of allowable SME scores, thereby permitting the SMEs to provide more precise evaluations.

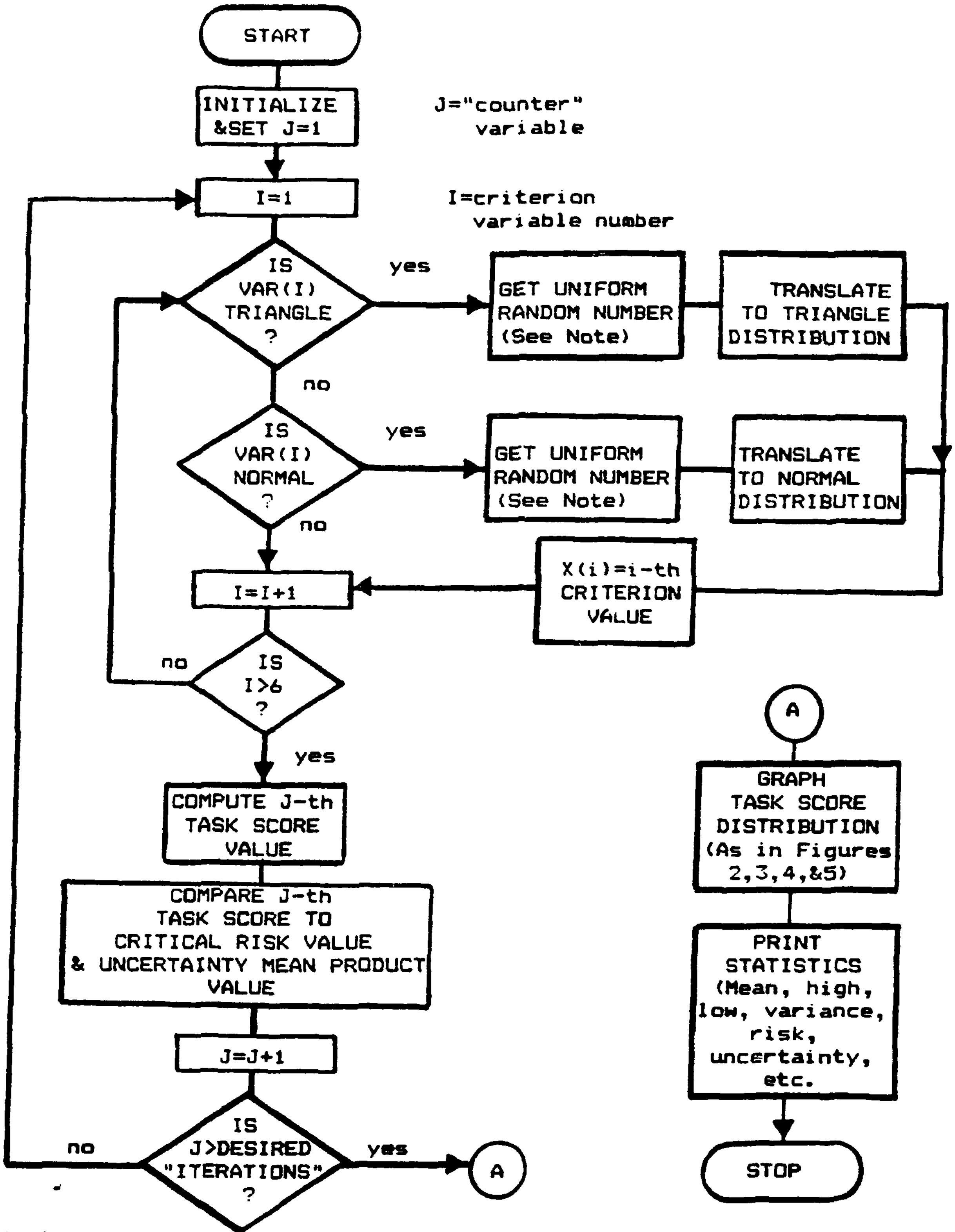
Another distribution that we could use is referred to as the 'triangle' distribution, which is characterized by the pessimistic (low), most likely (average), and optimistic (high) scores. In the triangle distribution, these three parameters are readily determined using SME sample data and the distribution requires no 'truncation.' The implications on the choice of either the normal or triangle distribution will be discussed later.

Other types of distributions may also be considered in the risk/uncertainty analysis, and the selection of the type of distribution certainly should not be based on 'the easiest to use' criterion. To help make an appropriate selection, the SME data should be plotted to see if it resembles the proposed distribution. SME responses need not be normally distributed (or exhibit any particular type of distribution) to conduct a risk/uncertainty analysis. Use of normal and/or triangle distribution herein is a simple matter of convenience.

Given our acceptance of the six criteria and the ability to use SME sample data to construct a distribution of all possible SME scores (using the SME sample to estimate the population parameters), we can now construct a theoretical population Task Score by randomly sampling from each of the six distributions and calculating the product of the randomly sampled values. The process of constructing the distribution of Task Scores is illustrated in Figure 1.

We note in Figure 1 that there is a 'counter', which is 'iterated' or increased as we calculate Task Scores. Thus, each particular count is frequently referred to as an 'iteration.' To construct a 'good' distribution of Task Scores, several thousand iterations will normally be required, thus necessitating a computer to automate the process. If we denote each randomly selected (triangle or normal) criterion variable as X_i (i = 1,2,3,4,5, or 6), then the j-th Task Score value is computed as:

Task $\text{Score}_j = X_1X_2X_3X_4X_5X_6$ (the product of criterion scores).



Note: A Uniformly distributed random number has equally probable numbers in the range of 0.0 to 1.0. See text reference (3), Section 8.

Figure 1. Generating the Task Score Distribution

We are now prepared to assess the uncertainty associated with the task using the procedure illustrated in Figure 1. First, we compute the Mean Product Task Score, which is simply the product of the means of the six criterion distributions (we know the means since we calculated them to describe the sample distribution parameters for both the normal or triangle distributions). The Uncertainty area lies to the right of the Mean Product Task Score as shown in Figure 2.

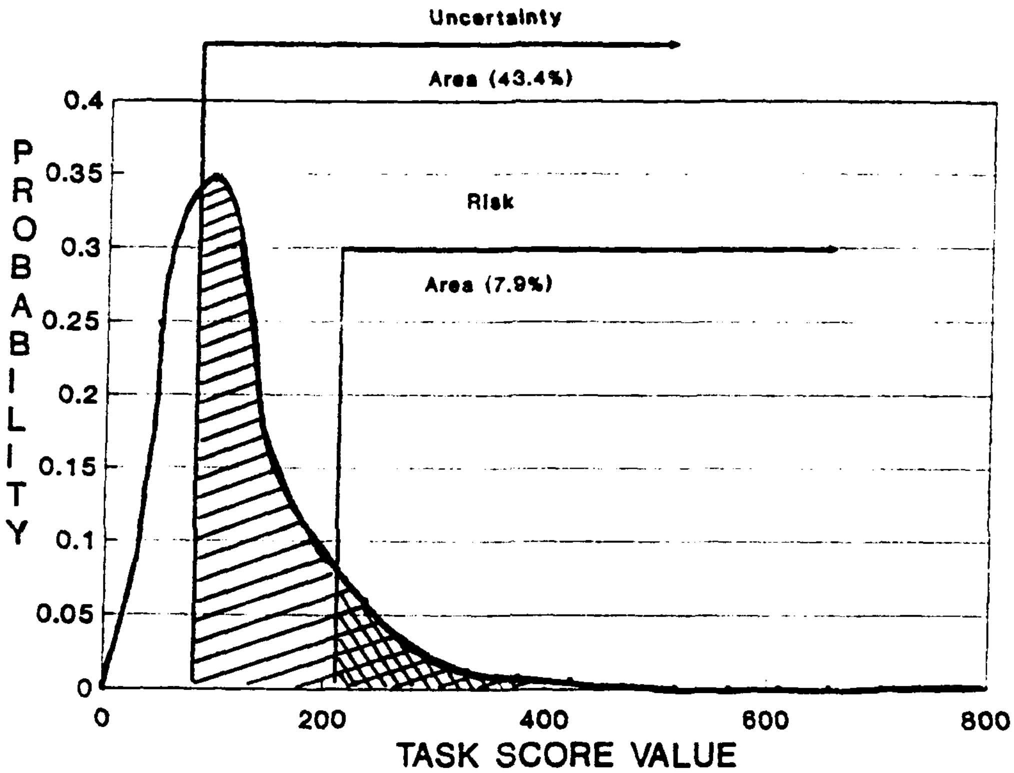


Figure 2. Risk/Uncertainty Using Normal Criteria Distribution

The Uncertainty is most often expressed in terms of a probability. We can estimate the uncertainty probability by counting the number of times the Task Score exceeds the Mean Product Task Score. (Recall that the ECA methodology computed the product of the means of the six criterion scores to compare to the cutoff score.) The probability is then calculated by dividing this count by the number of iterations or 'counter' value. For the results shown in Figure 2, there were 434 times that the Task Scores exceeded the Mean Product (87.965) out of the 1,000 iterations, so the uncertainty probability for this task example is 0.434.

The uncertainty probability tells us that there is less than an even chance (or probability less than 50 percent) that some other sample of SMEs could have given us a Task Score greater than the Mean Product Task Score. The uncertainty probability for the Task Score distribution is a composite of the uncertainty (disagreement) in the six criterion distributions. The value of uncertainty, then, is entirely a result of this disagreement relative to an average composite (Mean Product Task Score).

As we observed earlier, the ECA methodology compares the Task Score to an established cutoff score of 216. Now, however, we have a Task Score distribution to compare to the established cutoff score. The risk in the task is computed as a probability using a procedure similar to that which calculated the uncertainty probability. That is, we count each time that a Task Score value exceeds the established cutoff value. The risk probability is calculated by dividing that count by the number of iterations. For the example in Figure 2, the risk probability is 0.079, a moderately low value (some would say that a risk of 5 percent or more is 'significant').

What the risk probability is telling us is that despite a large uncertainty (probability = 0.434), the likelihood that the SMEs (or another sample of SMEs) would exceed the established cutoff value is only 79 times out of 1,000 (probability = 0.079).

At this point, we must somewhat disassociate the uncertainty from risk. In another situation, our SMEs could have absolutely agreed on a rating for a task (uncertainty equals zero percent), but the Task Score could have exceeded the critical value (risk equals 100 percent). Also, as we will see later, it is entirely possible for the risk probability to exceed the uncertainty probability. Thus, uncertainty does not necessarily cause or lead to risk.

So far, using this methodology, we can provide management with not only the 'standard' ECA methodology information, but we can also quantify the risk and uncertainty as well. As we will see, however, we can provide even more information to a decision maker.

Perhaps the easiest enhancement to make (with the intention to provide more information) is to change the established cutoff scores to 'critical' values which are 'tailored' to the situation being evaluated. From Figure 2, it can be seen that raising or lowering the critical value will affect the value of the risk. The value 216 (the ECA methodology established cutoff value) is arbitrary and may be considered to be too low or too high, depending on the situation being evaluated. Also, we compute a particular Task Score by forming the product of individual criterion scores. By working 'backwards,' we can conclude that each criterion has a cutoff imputed which is the sixth-root of 216, or roughly 2.45 (that is, 2.45 to the sixth power is about 216). This means that we have established that any score for any criterion above 2.45 is a 'risk' indicator. As we sampled from the six criteria and formed our distribution of Task Scores, we added up the counts of those Task Scores which consistently exceeded the critical value, thereby giving us a 'risk' assessment.

4. IMPORTANCE AND WEIGHTS.

At this point, we might make an observation that we may not consider each criterion to have equal importance in determining risk. The importance issue can be explicitly addressed by 'weighting' the criteria differently. Assume for the moment that we obtained the following 'importance' weights from our panel of SMEs using a procedure which we will explain below:

Table 3: Example of Importance Weights.

Distribution	Relative Importance	Normalized Weights
PP	10	1.50
TLD	5	0.75
TPD	1	0.15
FR	18	2.70
DR	2	0.30
TT	4	0.60
TOTAL	40	6.00

In the column labeled 'Relative Importance' in Table 3, FR is not only the most important criterion, it is 18 times the importance of TPD, as determined by our SME sample using the procedure which will be discussed below. These values of 'Relative Importance' must be transformed into weights, where the values of the weights must have certain properties.

First, since we have six criteria, the sum of the weights must equal six. This is needed to enable us to compare 'weighted' results to 'unweighted' results. (Actually, unweighted results have an imputed relative importance for each criteria of 1, so the sum of the relative importance equals the number of criteria). We can transform or 'normalize' the relative weights by dividing each of the relative weights by their total and multiplying by the value 6. The normalized weights are in the right column of Table 3. (For example, the weight for PP is found by dividing the Relative Importance by 40 and multiplying the result by 6). We see that the ratio of the weights is the same as the ratio of the relative importance. For example, the PP to TPD ratio is 10/1 for the relative importance and 1.50 to 0.15 for the weight. The ratio of importance has been preserved so that the 1.50 weight is still 10 times the 0.15 TPD weight.

To perform a risk/uncertainty analysis using weighted (importance) criteria and Task Scores, the calculation procedure is changed somewhat as follows:

$$\text{Task Score} = X_1^{wt_1} \times X_2^{wt_2} \times \ldots \times X_6^{wt_6},$$

$$\text{Risk Value} = CV_1^{wt_1} \times CV_2^{wt_2} \times \ldots CV_6^{wt_6}, \text{ and}$$

$$\text{Mean Product Test Score} = \bar{X}_1^{wt_1} \times \bar{X}_2^{wt_2} \times \ldots \bar{X}_6^{wt_6},$$

Where X_i is a randomly selected score from the i-th criterion distribution ($i = 1, 2, \ldots 6$ criteria),

CV_j is the critical value (constant) for the j-th criterion, and

$\bar{X}_j$ is the mean of the j-th criterion distribution.

Thus, the weights are exponents for the variables which are used for determining the Task Score distribution, critical value, and Mean Product Task Score. When each weight (wt_i) has a value of 1.0, the results are identical to the previous discussion results using 'unweighted' values.

There are several similar techniques which use the geometric mean for calculating weights. Care should be taken, however, to assure that the 'normalizing' process results in weights whose sum equals the number of criterion. Also, other techniques impose a

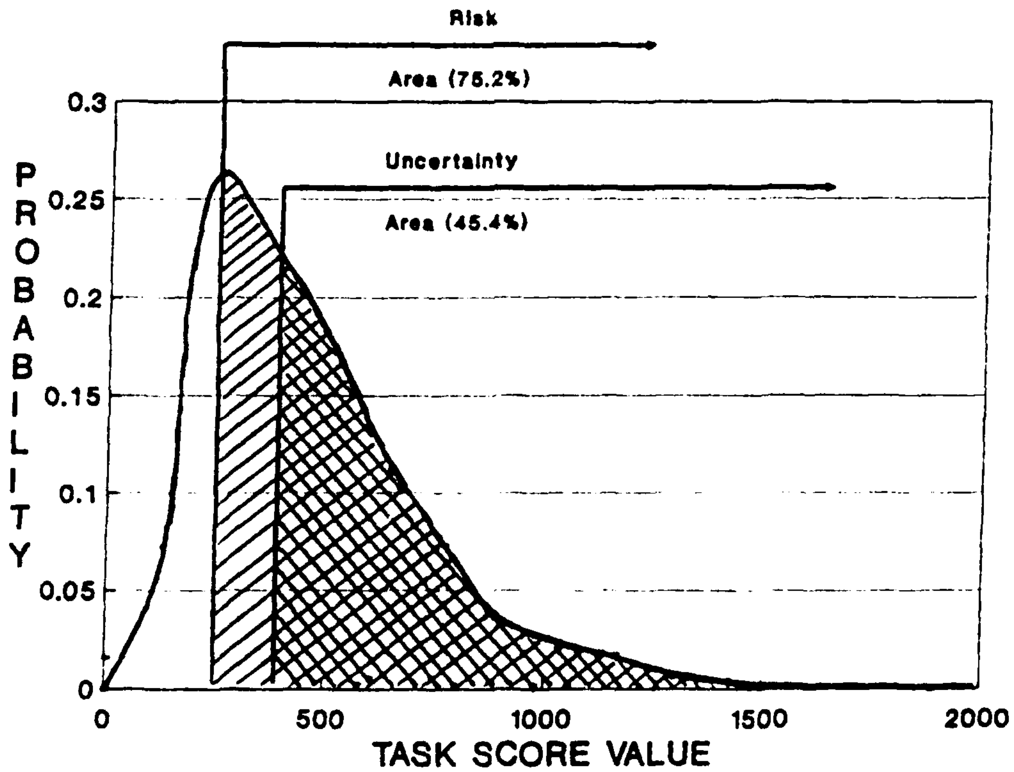


Figure 3. Weighted Uncertainty/Risk Using Normal Criteria Distributions

range of allowable scores (e.g., a fixed response scale) and require pair-wise comparisons of every criterion to every other criterion. Obviously, if a task is evaluated on a large number of criterion, the number of pair-wise comparisons could become large. See References c(3) and c(4) in Section 8.

Using the weighting procedure and the weights in Table 3, we obtain the Distribution of Task Scores shown in Figure 3.

The only change in the method used to generate Figure 3 and Figure 2 is that Figure 2 had equal criteria importance (weights) and Figure 3 used unequal importance (weights in Table 3). Note, however, there is a somewhat modest increase in Uncertainty, but a very large risk probability of 75.2 percent! Admittedly, the importance values were selected to give such a dramatic change in risk to illustrate a point. From Table 3, PP and FR had the largest weights <u>and</u> from Table 2, these also had the highest averages. This means that the contribution to the Task Score distribution due to these two criteria was far more than the contribution due to the other four criteria. Obviously, then, high criterion scores translate into high Task Scores, which translated into high risk. The point being illustrated is that risk is influenced by the importance placed on the criteria and is not influenced by the uncertainty of the SMEs.

Using Table 4, we can illustrate one method for calculating importance weights.

Table 4. Relative Weight Example.

	CRITERION					
SME	PP	TLD	TPD *	FR	DR	TT
1	9.0	4.0	1.0	11.0	3.0	7.0
2	10.0	7.0	1.0	20.0	1.5	2.5
3	16.0	5.0	1.0	25.0	0.5	3.0
4	8.0	4.0	1.0	18.0	3.6	3.0
5	9.0	6.0	1.0	19.0	2.0	6.5
Geometric Mean **	10.0	5.0	1.0	18.0	2.0	4.0

* Arbitrarily set to 1.0 as 'reference'
** Rounded to a whole number

For PP, the Geometric Mean = $\sqrt[5]{PP_1 \times PP_2 \times PP_3 \times .. PP_5}$

A similar calculation is used for other 5 criterion geometric means.

In Table 4, SME 1 felt that PP was nine times as important as (relative to) criterion TPD, which was given a reference value of 1.0. Thus, each SME rates each criterion importance relative to the reference criterion, so the values are relative importance (relative to the 'reference'). We note here, as in the SME evaluations in Table 2, that the SME have some uncertainty in their assessment of relative importance. The Geometric Mean is simply the 5-th root of the product of the SME scores for each criterion. The 5-th root is used in this case, since we have five SMEs. If we have 'n' SMEs, we would take the n-th root of the product of the SME relative importance values for each of the six criterion. It may occur that several SMEs would consistently rate a criterion lower than 1.0 (the 'reference'). We can see that SME 3 rated DR at 0.5 (meaning 'lower' or 1/2 as important as the reference). If the SMEs consistently rate a criterion at less than the 'reference,' the Geometric Mean will also be lower than 1.0. We saw in Table 3 how the Relative Weights (Geometric Means) are converted into 'normalized' weights. We should also observe that no negative or zero values are allowed as SME relative weights since such an evaluation has no practical meaning. (If the SME feels that the importance of a criterion is very small, a value such as 10^{-6} could be given, but not zero. Any zero value will make the geometric mean equal to zero regardless of all of the other scores. The 'dampening' effect caused by one SME providing very small numbers decreases rapidly as the number of SMEs (e.g., the sample size) increases to the 20-30 range.

As with the calculation of the Task Scores (which is similar to the Geometric Mean in that it involves taking the product of numbers), the use of the geometric mean for calculating the importance weights requires several assumptions to be met and, as will be shown, 'builds' certain properties into the values thus calculated.

First, we must assume that the SMEs rank the relative importance of each criterion independently of each other and that the ranking of one criterion does not affect (or is not influenced by) another criterion (other than the 'reference' criterion). Secondly, we must also assume that the values given by the SMEs are samples of all possible SME values. With these two assumptions, we can use the properties of the SME relative importance distribution to infer something about the importance of the criterion if we could get all SMEs (vs. a sample only) to rate the criterion.

However, unlike our sampling procedures used to build the Task Score Distribution, we need a set of specific weights to use. Yet, we want certain statistical properties to be present when we calculate these weights. The two desired statistical properties are that the procedure should capture both the overall (concensus) value and be sensitive to wide variations in SME ratings (e.g., uncertainty or disagreement). Thus, the selection of the techniques for calculating the importance weights is not arbitrary: it must possess certain properties. The data in Table 5 will be used to illustrate the effect of disagreement of SMEs on several commonly used statistics.

Table 5. Illustration of Statistical Properties

SME	Alternatives 1	2	3
1	4.50	3.00	7.00
2	5.00	3.80	1.00
3	3.00	4.60	7.00
4	2.00	3.00	4.00
5	3.00	4.60	1.00
Mean	3.50	3.80	4.00
Variance	1.20	0.51	7.20
Geometric Mean	3.32	3.73	2.87

Illustrated in Table 5 are the scores by five SMEs on three alternatives. The particular interpretation of the meaning of the alternative or score is not important. If we were to select an alternative with the highest mean (average), we would select Alternative 3, with a mean of 4.0. Note, however, that Alternative 3 had the most 'disagreement' among the SMEs as evidenced by the highest variance. On the other hand, Alternative 2 has a somewhat lower mean (3.80) and has the lowest variance (hence, the most agreement by the SMEs).

The geometric mean has the property that the highest value is achieved when the variance is the lowest when the mean remains constant. For example, try any three positive (not zero) numbers that add to some value. Change the numbers at will so long as their sum remains the same. We will find that the geometric mean of the three numbers will always be less than or equal to the arithmetic mean (the average) and will equal the arithmetic mean only when all three numbers are identical (e.g., the variance is zero).

The implication of this is that the geometric mean combines some properties of both an ordinary average and a variance. Using this technique for calculating the relative weights 'builds' in several desirable statistical properties. First, if one relative weight is greater than another (is more important), it must have consistently higher scores <u>and</u> the SMEs must agree that that it is an important criterion (the variance is low). Conversely, a particular criterion may have a low weight (importance) either because all of the SMEs uniformly rated it low or the criterion had a large variance in score (the SMEs disagreed on its importance).

5. <u>TASK CRITERIA DISTRIBUTIONS</u>.

Thus far, we have not discussed the implications of the choice between a normal or triangle distribution for the six criteria. Figures 2 and 3 respectively showed the unweighted and weighted distribution of Task Scores using a normal distribution for all six criteria (the distributions differ, however, in their mean and variance, but look alike in that they are normally distributed). In the unweighted case, we found a small but significant risk probability of 0.079. Figures 4 and 5 show the unweighted and weighted Task Scores, respectively, using all triangle distributions for the six criteria (each has the lowest SME value, the average, and the highest SME value as its parameters).

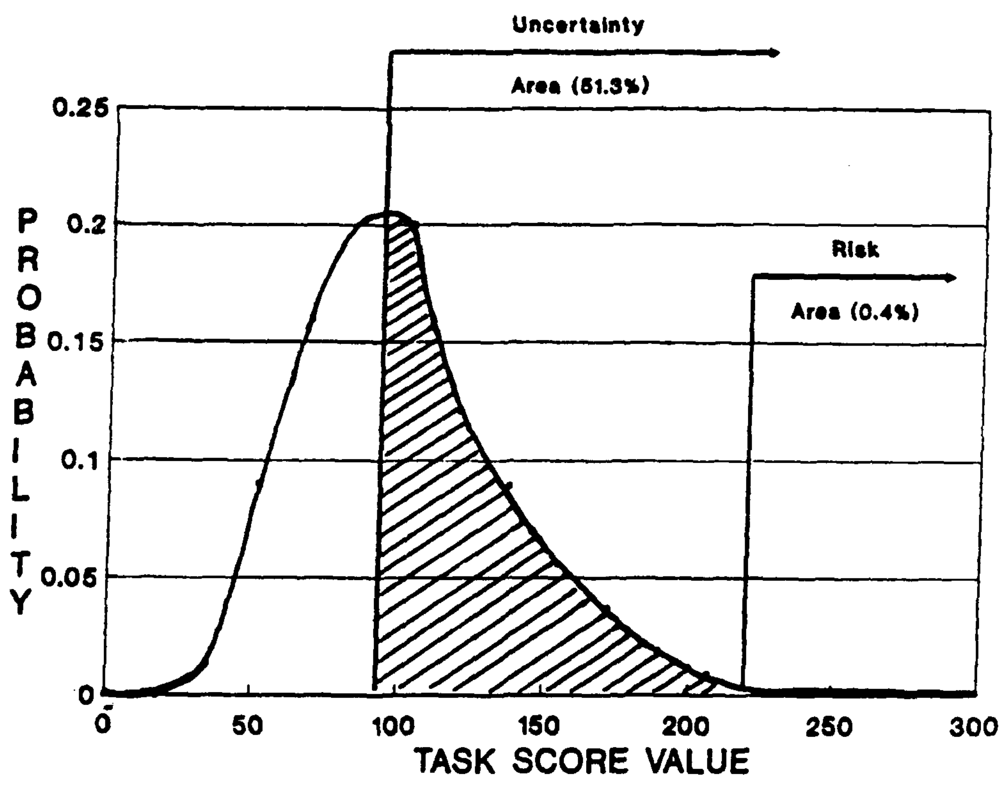


Figure 4. Unweighted Task Scores Using Triangle Distributions

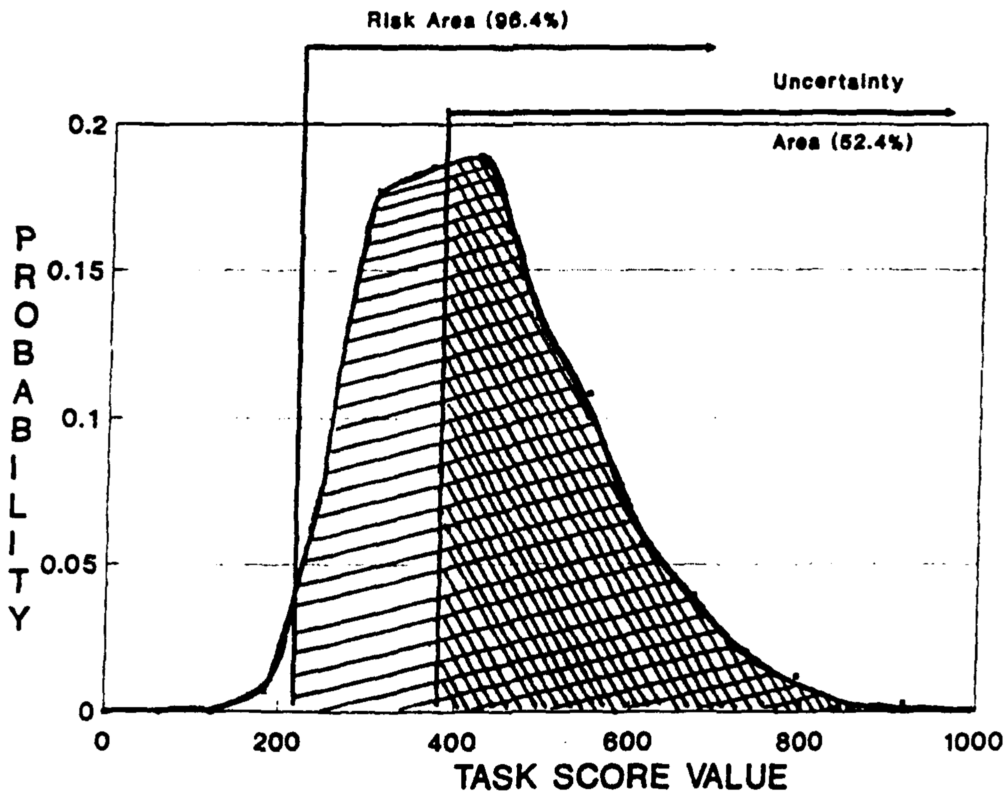


Figure 5. Weighted Task Scores Using Triangle Distributions

From Figure 4, we can see that the risk probability decreased substantially (0.4 percent), while the uncertainty probability increased somewhat (51.3 percent) relative to the results in Figure 2. In this situation, this is to be expected since the triangle distribution 'cuts off' some of the values which the normal distribution allows (even though the normal distribution is 'truncated'). For example, with a triangle distribution with low = 3.0, medium = 3.4, and high = 4.0, there will be no criterion scores generated lower than 3.0. However, a normal distribution with an average equal to the triangle distribution's average (3.4) can have values lower than 1 or higher than 4 were it not for the 'truncation' which we imposed (note, truncation takes place because we imposed a limit of a score from one to four on the SME's choice, and therefore, we impose this on allowable criterion scores as well).

The triangle distribution with a 'high' (optimistic) parameter of 2.0 can generate values no higher than 2.0. However, for the same criterion with a normal distribution with mean of 1.4 and standard deviation of 0.55, random numbers up to two standard deviations should be generated (value of 2.5). Thus, the triangle distribution generates lower criterion values, meaning it gives a lower estimate of risk in this case. This change may not always be present when 'real' data and situations are used (for example, if a triangle distribution has a low value of 1 and a high value of 4, it will provide values as high and low as a normal distribution with a mean of 2.5 and variance of 1.0. However, the frequency of high and low values will differ between the normal and triangle distribution.

Note, however, that the risk in Figure 5 has increased from the risk in Figure 3 even though the only change is in the distribution types for the criterion variables. Whereas Figure 3 used the same weights used in Figure 5, the criterion distributions changed from normal to triangle and the risk increased because of the weighting and the type of distribution of the criterion variables.

The practical implication of this discussion is that the choice of the scales for SMEs to rate is critical (in this discussion, the allowable range was 1 to 4 to reflect the ECA methodology). When the range of the scale is increased, we will most likely see more variation in responses, but the risk/uncertainty methodology can accommodate this much better than the ECM methodology. In fact, to get a better risk/uncertainty assessment, the range of allowable scores for SMEs should be widened to the maximum extent possible.

6. <u>OVERALL RISK</u>.

Thus far, we have explicitly addressed the risk/uncertainty of a single task. Assume, for purposes of illustration, that the proposed system has ten tasks that are independent (that is the completion of one task is not dependent on and does not affect the completion of another task). By applying the risk/uncertainty analysis as previously discussed, we can obtain the risk/uncertainty of each individual task. Through probability theory, as shown below, we can determine the overall risk for the proposed system. This, of course, implies that we could select one alternative configuration over another by computing the alternative risks and selecting the lowest risk alternative. Additionally, if the overall risk is too high, we can 'backtrack' through the analysis to find the high risk/uncertainty tasks/criteria and single these areas out for further study. Study, would then be used to reduce the uncertainty/risk by 'getting a handle' on what the source of the high risk is.

Table 6 illustrates the risk assessments of six independent tasks for a proposed system. We should note that these are individual tasks, and are not the criteria used to evaluate a task as shown in Table 2. We can calculate the overall risk by taking the product of the 'unrisk' (that is, one minus the value of the risk) and subtracting the result from 1 as shown below:

Table 6. Multitask Risk Assessment

Task	Risk	1-Risk
1	0.20	0.80
2	0.10	0.90
3	0.01	0.99
4	0.30	0.70
5	0.01	0.99
6	0.35	0.65

$$\text{Overall Risk} = 1-[(.80) \times (.90) \times (.99) \times (.70) \times (.99) \times (.65)]$$

$$= 0.679.$$

Overall, there is a high risk for this system (probability = 0.679). By 'backtracking,' we can see that there are three tasks (1, 4, and 6) that are significant factors in the overall risk determination. Further 'backtracking' can be done to point to specific risk areas (the criterion) and provide management with a means of selectively using further study and analysis to reduce risk/uncertainty and to identify issues needing more intensive investigations.

If the tasks are not independent, we should not use the product form of overall risk. A computer model of the analysis could be developed to assess the risk if the type of task dependency is known. Also, some tasks may be more important than others (for example, some task may cost more than others), so an overall risk determination can be made using task weights and a methodology similar to the weighting of criterion distributions. There would be a variety of methods which could be used (including SME opinion) to determine appropriate task weights.

7. CONCLUSIONS.

The use of techniques which reduce a set of data to only an average do not permit us to use all of the information which might be obtained from our Subject Matter Experts (SMEs). By explicitly considering the SME disagreement, we can determine uncertainty, and by comparing the calculated Task Scores to an appropriate critical value, we can calculate the risk in proposed materiel acquisition systems.

This uncertainty and risk analysis can be used to guide the needs for further analysis to reduce the risk that critical tasks will become 'high drivers' in the acquisition process. SMEs can also provide valuable information on the importance of individual criteria used to evaluate the proposed system, thereby directly influencing the determination of the risk as well.

We have seen that a high uncertainty does not necessarily imply a high risk. Uncertainty has to do with the amount of disagreement of our experts and risk has to do with the likelihood that some critical consideration will exceed some predetermined value.

We have also observed that the choice of scales that the SMEs use in evaluating the criteria will definitely affect the risk/uncertainty determination since values will be 'truncated' when we sample from the SME criteria distributions. Accordingly, when performing a risk/uncertainty ECA, the range of allowable SME responses should be widened as far as possible so as not to restrict unnaturally the choice of SME scores (thereby decreasing our ability to assess their disagreement and hence, uncertainty).

8. REFERENCES.

a. Army Regulations:

(1) AR 40-10, Health Hazards Assessments.

(2) AR 71-9, Materiel Objectives and Requirements.

(3) AR 385-16, System Safety Engineering Program.

(4) AR 602-1, Human Factors Engineering Program.

(5) AR 602-2, Manpower and Personnel Integration (MANPRINT) in the Materiel Acquisition Process.

b. Other Army Publications:

(1) MANPRINT RISK ASSESSMENT procedures guide, SSC-NCR *, ATTN: ATNC-NMF-AIB, 200 Stovall Street, Alexandria, VA 22332-0400, September, 1987.

---

* SSC-NCR is now called the US Army Personnel Integration Command (USAPIC).

(2) Early Comparability Analysis Handbook, SSC-NCR.

(3) MANPRINT Analysis Methodology (draft), Directorate for MANPRINT Research and Studies, Office of the Deputy Chief of Staff for Personnel (ODCSPER), Washington, DC.

(4) HARDMAN Comparability Analysis Methodology Guide, US Army Research Institute (ARI) and SSC-NCR.

(5) Memorandum, ATCD-ET, dated 31 May 88, Subject: Concept Formulation Process (CFP) Memorandum of Instruction (MOI).

c. Text Books:

(1) Bartle, Robert G., The Elements of Real Analysis, John Wiley & Sons, Inc., 1964.

(2) Dalkey, Norman C., 'Delphi,' RAND, P-3704, October, 1967.

(3) Hwang, Ching-Lai and Abu Syed Md., Masud, Multiple Objective Decision Making-Methods and Applications, Lecture Notes in Economics and Mathematical Systems (#164), Springer-Verlag, 1979.

(4) Hwang, Ching-Lai and Abu Syed Md., Masud, Multiple Attribute Decision Making, Lecture Notes in Economics and Mathematical Systems (#186), Springer-Verlag, 1981.

(5) Law, Averill M., and W. David Kelton, Simulation Modeling and Analysis, McGraw-Hill Book Company (Series in Industrial Engineering and Management Science), 1982.

(6) Lindgren, Bernard, Statistical Theory (Third Edition), MacMillan Publishing Co., Inc., 1976.

(7) Lodge, Milton, Magnitude Scaling: Quantitative Measurement of Opinions, Book 25 from the series: Quantitative Applications in the Social Sciences, a SAGE University Paper, 1981.

(8) Luce, Duncan R., and Raiffa, Howard, Games and Decisions, John Wiley & Sons, Inc., 1967.

(9) Phillips, Don T., et al, Operations Research - Principles and Practice, John Wiley & Sons, Inc., 1976.

9. A POST SCRIPT.

The objective of this paper is to be provocative. Admittedly, there are many considerations which have not been addressed. The purpose of the paper is to illustrate that existing methodology can be expanded when proper consideration is given to the analytical methodologies.

When the first drafts of the paper were distributed for comments, several discussion points emerged. Although the discussion below may not address all such possible considerations (the literature abounds in such considerations), it does address and illustrate the need for further work in the ECA methodology. The discussion below concludes with a recommendation on the development of an Expert System for use by non-technical people. Such a system would have built in features to appropriately incorporate the theoretical considerations discussed below.

Independence of criterion scores loomed as a major concern to several readers. If we were using the criterion scores to build an index or some other scale which could be used to estimate or predict a proposed weapon system's performance, we would want them to be independent in a statistical sense as well as independently evaluated by our SMEs. In the use of ECA methodology, there is evidence to suggest that Task Learning Difficulty (TLD) and Time-to-train (TT), for example, are <u>not</u> statistically independent. Other criteria (or subscores) may also be correlated. Therefore, when we multiply the six criterion variables to obtain the ECA score, one might argue that the score is biased because of the correlation between the subscores. For the ECA methodology however, the independence consideration is not an issue.

We are interested in <u>relative</u> measurements. Unlike an IQ, for example, which might be estimated by a series of subscores on a 'standardized' test, the ECA score has no intrinsic value in and of itself. The ECA score takes on importance only with respect to some arbitrarily established cutoff point which is a composite of the individual subscores themselves. Any 'bias' in the ECA score will be similarly present in the cutoff score.

If we insist on concerning ourselves with correlation between subscores or criterion variables, we might look at several possible causes of correlation. To a large extent, the reason for dependency rests in the use of a very restricted range of SME allowable scores. Certainly, widening the allowable range of scores (e.g., from a 1 to 4 range to a range of 1 to 7) gives the SME more choices. If SMEs use the increased range, we should see the 'dependency' reduced. Suppose, for example, that five SMEs rate a criterion which has a range of 1-4 for allowable scores. The first four can each select a different score, but the fifth <u>must</u> select a score which duplicates at least one of the previous four SME scores. Thus, the SME scores will become statistically dependent simply because we restricted the allowable range of scores. If we expand the range, we increase the allowable responses. Thus, an objective of increasing the allowable range is to increase the variance in the SME response (assuming, of course, that the SME will use the increased range). However, the wider allowable range of scores may increase the SMEs' difficulty in discerning more subtle differences in the ratings.

If, nonetheless, a reasonable goal is to increase variance, another variance-increasing technique actually is present in the ECA methodology. The product form of the ECA creates the maximum variance in the ECA score where there is little variance in the individual subscores. Recall that the risk/uncertainty method is essentially concerned with variance. Variance in SME opinion translates into uncertainty. Variance in ECA scores that are above a certain 'critical' value is a measurement of risk. Variance is caused by disagreement of the SME. Thus, we want a method which calculates the most variance from our sample of SMEs.

Note that a method which calculates an average score not only does not generate the maximum variance, it also allows for 'compensatory' ratings. That is, if one SME scores one point above the average and another scores one point below the average, for example, then the two SME have compensated for each other, thereby 'averaging out' their disagreement. The product form of the ECA allows for no such averaging out or compensation of SME scores and, in fact, allows us to focus on disagreement.

Another cause for apparent dependency may be that the criterion variables are, indeed, related. For example, if a task is difficult for the soldiers to learn, it probably is also difficult to train, and difficult training tasks most likely will take more time to train. Hence, TLD and TT might reasonably be expected to be related or dependent.

If, however, we wish to construct an ECA score which accounts for such dependency, we can simply weight the individual subscores by some appropriate value. For example, if TLD and TT were the only two correlated subscores, we could give each one a weight of one-half. Another technique for handling the correlated variables is to perform a factor analysis on the criterion variables to determine the minimum number of independent factors and the equation which transforms linear combinations of subscores into factor scores. The risk/uncertainty method can then be applied to the factor scores. We could also expand the number of subscores and use one of several available techniques (discussed above) to adjust for dependency if we feel this is necessary.

The point being made is that the lack of independence in our subscores is not a sufficient reason to discount the usefulness of the ECA methodology nor should it even necessarily be a concern. We are not necessarily after an absolute measurement scale. If we are concerned about correlated subscores, techniques exist to ameliorate perceived difficulty.

This paper obviously did not address a number of other issues concerning the use of the ECA methodology. For example, the issue of inter-rater reliability should be addressed in the adequacy of the design of the data collection instrument.

This paper also did not deal with how the results could be implemented. The analysis portion of the ECA risk/uncertainty methodology has been automated. We propose that an Expert System for ECA be developed as an Artificial Intelligence application. The system would include a processor which would allow the user to 'tailor' the questions for SME to rate and would have various consistency checks built in (i.e., internal rater consistency, inter-rater reliability, etc.). Obviously, such a capability will take more effort to complete, so this paper should be viewed as only a start in the process.

FOREWORD

The Early Comparability Analysis (ECA) methodology was developed by the US Army Personnel Integration Command (USAPIC) and was approved for use as a Manpower Personnel Integration (MANPRINT) methodology in February, 1985. It was designed as a lessons learned approach for analyzing those tasks performed by soldiers on currently fielded systems. In addition, it was designed to be easily performed by TRADOC action officers using resources that are readily available. The identification of 'high driver' tasks, those tasks which significantly affect Manpower, Personnel, and Training (MPT) resources, is completed by a relatively simple screening mechanism. The labor intensive work is to conduct a task analysis on the high driver tasks.

Many potential enhancements to the methodology, according to USAPIC, were deliberately omitted in the design of the methodology even though they may have been somewhat beneficial. The overall design goal was to keep the methodology easy to perform.

With modifiation, the basic ECA methodology can also be applied to conceptual systems early in the materiel acquisition process. This report demonstrates how risk and uncertainty measurement techniques can be applied to the basic ECA methodology to permit an early MPT assessment for a conceptual system. In addition, the techniques can be applied to fielded systems. The risk and uncertainty procedures provide a means of making the ECA effort to identify high drivers more rigorous if the resources to do so are available.